

Original citation:

Ng, Irene C. L. (2018) *The market for person-controlled personal data with the Hub-of-all-Things (HAT)*. Working Paper. Coventry: Warwick Manufacturing Group. WMG Service Systems Research Group Working Paper Series (01/18). (Unpublished)

Permanent WRAP URL:

<http://wrap.warwick.ac.uk/101708>

Copyright and reuse:

The Warwick Research Archive Portal (WRAP) makes this work by researchers of the University of Warwick available open access under the following conditions. Copyright © and all moral rights to the version of the paper presented here belong to the individual author(s) and/or other copyright owners. To the extent reasonable and practicable the material made available in WRAP has been checked for eligibility before being made available.

Copies of full items can be used for personal research or study, educational, or not-for-profit purposes without prior permission or charge. Provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way.

A note on versions:

The version presented here is a working paper or pre-print that may be later published elsewhere. If a published version is known of, the above WRAP URL will contain details on finding it.

For more information, please contact the WRAP Team at: wrap@warwick.ac.uk



**WMG Service Systems Research Group
Working Paper Series**

The Market for Person-controlled Personal Data with the Hub-of-all-Things (HAT)

Irene CL Ng

**ISSN: 2049-4297
Issue Number: 01/18**

About WMG Service Systems Group

The Service Systems research group at WMG works in collaboration with large organisations such as GlaxoSmithKline, Rolls-Royce, BAE Systems, IBM, Ministry of Defence as well as with SMEs researching into value constellations, new business models and value-creating service systems of people, product, service and technology.

The group conducts research that is capable of solving real problems in practice (ie. how and what do do), while also understanding theoretical abstractions from research (ie. why) so that the knowledge results in high-level publications necessary for its transfer across sector and industry. This approach ensures that the knowledge we create is relevant, impactful and grounded in research.

In particular, we pursue the knowledge of service systems for value co-creation that is replicable, scalable and transferable so that we can address some of the most difficult challenges faced by businesses, markets and society.

Research Streams

The WMG Service Systems research group conducts research that is capable of solving real problems in practice, and also to create theoretical abstractions from or research that is relevant and applicable across sector and industry, so that the impact of our research is substantial.

The group currently conducts research under six broad themes:

- Contextualisation
- Dematerialisation
- Market/Platform Design and Economic Models
- Service Design
- Value and Business Models
- Viable Service Systems and Transformation
- Visualisation

WMG Service Systems Research Group Working Paper Series

Issue number: 01/18

ISSN: 2049-4297

May 2018

The Market for Person-controlled Personal Data with the Hub-of-all-Things (HAT)

Ng, Irene CL

Service Systems Research Group
Warwick Manufacturing Group (WMG)
University of Warwick, Coventry, CV4 7AL, UK
Tel: +44 (0) 247652 4871
E-mail address: Irene.Ng@warwick.ac.uk

Acknowledgement

The authors gratefully acknowledge the funding contribution of the Research Council (UK) Digital Economy and the EPSRC to the Hub-of-all-Things (HAT) project (<http://hubofallthings.org>) through the following awards: Home Hub-of-all-Things (HAT) as Platform for Multi-sided Market powered by Internet-of-Things: Opportunities for New Economic & Business Model (grant reference EP/K039911/1); Smart Me versus Smart Things: The Development of a Personal Resource Planning (PRP) System through Human Interactions with Data Enabled by the IoT (grant reference: EP/L023911/1); ConTriVE -)Control and Trust as Moderating Mechanisms in addressing Vulnerability for the Design of Business and Economic Models (grant reference: EP/N028422/1); ACCEPT: Addressing Cybersecurity and Cybercrime via a co-Evolutionary aPproach to reducing human-relaTed risks (grant reference: EP/P011896/1).

If you wish to cite this paper, please use the following reference:

Ng, Irene C. L. (2018) *The market for person-controlled personal data with the Hub-of-all-Things (HAT)*. Working Paper. Coventry: Warwick Manufacturing Group. WMG Service Systems Research Group Working Paper Series (01/18). (Unpublished)

The Market for Person-controlled Personal Data with the Hub-of-all-Things (HAT)

Irene CL Ng

Abstract

This paper theorises personal data more fundamentally and exposes the vulnerability of organisation-controlled personal data. It then reports on the micro-economic model design of a new personal data exchange by designing and creating PPD or person-controlled personal data through a new technological artefact called the Hub-of-all-Things (HAT) and report the build and implementation of the mechanisms, coordination and incentive structures for PPD exchanges.

Keywords

Personal data, markets, micro-economic model, privacy, game theory, signaling

Introduction

Data is an institutional artefact. It is a symbol and a visual cue, making sense only because of our beliefs. Early institutional economists consider such beliefs as an institutionalised logic, a form of social structure that shapes behaviours, purposes or preferences (Commons 1931). I can conceive the ‘first’ data to be scratches on a cave wall, markings to denote the height of a child. If one did not know it was there for this purpose, it would just be scratches on the wall. However, by the nature of the way the scratches are systematically carved, I believe they could be height markings. Hence, I rely on the institutions I have crafted to make sense of the scratches.

When data becomes digital, it is encoded as a *bitstring*, a sequence of binary digits, 0s and 1s (Quah 2003). Economists have studied bitstrings as digital goods (Quah 2003; Belleflamme 2016) or information goods beginning from Shapiro and Varian’s (1998) treatise that anything that can be digitised is an information or digital good in the economy. To make sense of such bitstrings, certain assumptions are made; chief amongst them is that the bitstring is valuable i.e. it affects the payoff to some actor in the economy, which is why it was created in the first place.

This paper builds upon work on digital goods, first by theorising the economic properties of personal data and its derivative, its inferential signals that are of value to firms in the market. From there, I expose the vulnerabilities of personal data in control of firms and the market failure of perishable signals, signals generated by personal data that expire quickly. I then present a redesign of the market for personal data through the creation of a new asset; person-controlled personal data (PPD), a new asset class arising from individual ownership of a personal micro-

server that bestows intellectual property rights over the personal data within it. I elaborate on the microeconomic design and build - the asset, the coordination structure and the market design for PPD. Finally, I report the findings from the market implementation.

Personal Data as an Information Good

Personal data is generated from the consumption of goods and services. For example, interacting on social media generates posts; using Google Calendar creates entries. Even consuming physical goods generates personal data, from wearing a Fitbit to using an Internet-connected coffee machine. The bitstring data collected is usually stored within software applications and forms the asset of the firm that owns the app and the technology. Such personal data may not be released by the firms that own it, but should the firm transfer it on, it reaches the market in two ways.

The conventional route is its release to data brokers. This release is sometimes combined with de-identification, e.g. scrubbing personal identity information from the rest of the data, replacing it with a generic ID, especially in parts of the world with strong data protection and privacy laws. Since the advent of ICT, consumer data in the form of supermarket spending and banking transactions have been collected, analysed and often transferred to third parties and re-combined with other data. Often, the anonymised data is “re-identified” so as to draw insights that help brands plan their consumer strategies. Re-identification is the practice of matching de-identified data with other datasets in order to discover the individual to whom the data belongs. It is through this process that firms have been able to target their offerings to the consumer segment most likely to purchase their products. Such consumer data link actual behaviours gleaned from disparate datasets to predict future buying and usage behaviours and sell on the insights to firms.

The advancement of Internet technology has resulted in a new channel for transferring personal data; a set of clearly-defined standards of communication between software components (whatever software language they are written in), called Application Protocol Interfaces (APIs). APIs are now one of the most common ways technology companies integrate with one other. An example would be the sharing of Spotify (music streaming) data with Sonos (speaker) resulting in individuals being able to play their own Spotify playlists on Sonos speakers. Sharing personal data in this manner results in contracts between firms for data usage, with the user’s consent, and moves data from being a resource within ICT systems to becoming an information/digital good provided by a source firm as an input factor to the destination firm. The mutual sharing of personal data between applications with the consent of the individual achieves pareto-efficient outcomes since individuals benefit from re-using data that is locked-up within other applications, and all firms benefit from more data at low marginal costs. This liberation of personal data from firms to become a real-time, on-demand and dynamically-

updated information good is creating network effects but require new ways of conceptualising information goods in economics.

Take, for example, a person relaxing at the end of the day by making a cocktail, searching for the recipe online and then listening to music on Spotify while drinking the cocktail. While searching for the recipe is a consumption of a digital good (a browser), generating the data when searching or while listening to music is an informational resource to aid the improvement of Google and Spotify, similar to data generated by a car engine informing its future design and production, regardless of who drives it (Loebbecke 2002). Yet, that data is evidence of actual human behaviour, much like scanner data at supermarkets or transaction data in banks. When liberated, it contributes to the existing market for consumer data. Similarly, physical things with sensors becoming Internet-Connected-Objects (ICO) are starting to generate petabytes of data from room temperature, train arrival locations, doors or windows opened (Ng and Wakenshaw 2017). The creation of sensor data from human interactions is exploding the supply of personal data as an information good. This is especially so when personal data held in walled siloes of applications would generate higher payoffs when combined with other datasets (see Ng et al. 2015). The large amount of data generated and available for recombination has created the challenge of ‘big data’, described as data with ‘volume, variety and velocity’ (TechAmerica Foundation’s Federal Big Data Commission 2012; Gandomi and Haider 2015; De Mauro, Greco, and Grimaldi 2016). The big data movement often studies three overarching issues: (1) technology problems, ie. collecting, storing and analysing the large volume of data; (2) commercial value ie. insights from data; and (3) societal impacts of big data, ie. privacy, regulations for commercial use of this data (Nunan and Di Domenico 2013). Whether legal or not, personal data now accounts for 36% of data-brokering activities globally (Transparency Market Research 2017). When personal data becomes liberated by the firm through connectivity or re-selling, a secondary market emerges due to its potential benefit as an asset, particularly to advertisers and manufacturers, since it can be used to generate consumer insights albeit at some societal cost in terms of privacy loss (Acquisti, Taylor, and Wagman 2016). Leading companies operating in the global data broker market include Acxiom Corporation, Experian Plc, Equifax, Inc. Within these markets, data brokers aggregate and analyse consumers’ data to make inferences about them.

There is evidence that despite the high transaction costs and risks in holding personal data, re-identification of data through the connection of disparate datasets finds an efficient market. Thus, even if anonymisation and aggregation may destroy the original structure of the bitstring dataset, necessary to comply with some regulatory authorities like Europe, a market emerges to reconstitute it. Indeed, as a Federal Trade Commission study¹ has revealed, data brokers collect personal data from many resources largely without consumers’ knowledge and re-

¹ Ramirez, Edith, Julie Brill, Maureen K. Ohlhausen, Joshua Wright, Terrell McSweeney, (2014) Data Brokers: A call for transparency and accountability, Federal Trade Commission report

identify them for the purpose of increasing the value of the insight for sale to advertisers and marketers.

Research on personal data-sharing in the economics of privacy (e.g. Acquisti 2010 etc.) have found that disclosing personal data do bring benefits to individuals, such as immediate monetary compensation (e.g. discounts), intangible benefits (personalisation and customisation of information content) and price reduction as an effect of more targeted advertising and marketing, information-based price discrimination, and more targeted ads to better inform consumers (cf. Akcira and Srinivasan 2005). However, such sharing also brings about costs and negative externalities for example, privacy costs, and subjective and objective privacy harms. Conversely, it has also been suggested that sweeping privacy regulation that result in firms not being able to obtain personal data will lead to opportunity cost and inefficiencies (Acquisti 2010; August and Tunca 2006; Van Zandt 2004; Anderson and de Palma 2005; Hann et al. 2006).

With the increasing economic value of personal data, scholars have been polarised into two main camps. The first, regulatory camp advocates for privacy protection as an end in itself, regardless of economic consequences. The underlying notion of such an advocacy is that privacy is a human right to personal data protection. This is consistent with the EU Charter that data being processed for specified purposes and with consent of the person concerned or with some other legitimate basis is laid down by law (Godel, Litchfield, and Mantovani 2012, p.42; Charter of Fundamental Rights of the European Union, *OJ C 364*, p. 10, 18.12.2000, Article 8). Enforcement of regulation would also pose a challenge since there is doubt as to how much regulatory powers governments actually have over the Internet. Any attempt of territorial governments to enforce privacy regulations could increase the likelihood of data-driven companies (whose profits depend significantly on data) to employ legal arbitrage, moving to jurisdictions outside the regulation. In the extreme, adverse selection could drive out firms benefiting from the data economic chain, reducing tax revenues for the country.

With the continuing advancement of digital technology, the argument for personal data protection has evolved from the human-rights concern to an economic rationalisation based on the trade-offs between risks and return (Godel, Litchfield, and Mantovani 2012). This is the approach taken by the self-regulatory camp, proposing that a market solution exists as a trade-off between privacy and the benefit from data usage (Acquisti 2010). This camp proposes that individuals could be assigned property rights to the information so that they are able to contract with third parties on how they might use it. Legal scholars have advocated the 'propertization' of personal data and argued against the imposition of legal limits on data trade i.e. that there is no need for "inalienabilities" (i.e. any restriction[s] on the transferability, ownership or use of an entitlement (Rose-Ackerman 1985). The self-regulatory framework advocates the exchange of data and data protection to increase aggregate welfare, emphasising market self-correction for efficiency outcomes and the regulators' role as one of steering the market through a combination of incentives, disclosure policies and even liability (Acquisti 2010).

Unfortunately, the practical implementation of a self-regulatory framework faces huge challenges because many of the data exchange contracts are incomplete and there is very little transparency about the secondary uses for the data (Beresford, Kübler, and Preibusch 2010; Godel, Litchfield, and Mantovani 2012). Property rights are a challenge for individuals to exercise when the personal data is held by firms collecting the data and not by individuals themselves (Shapiro and Varian, 1997; Laudon 1996). Since personal data is often mixed with other data belonging to the firm, the lack of boundaries would make property rights for individuals too much of a challenge to implement and enforce, leading to higher transaction costs. In addition, third parties buying and selling personal data could impose social costs on individuals since individuals are not directly involved in these transactions, resulting in the externalities that are not internalised by the firm (Godel, Litchfield, and Mantovani 2012; Odlyzko 2003; Swire and Litan 1998; Acquisti 2010).

Aside from the challenge in implementation, others have also argued against the trade or propertization of personal data due to its impact on privacy. With the development of technology and devices that can generate finely-grained information about consumers' privacy preferences (McGeveran 2001), trade of personal data could lead to its commodification and contribute to additional privacy intrusions (Tuan, 2000). Additionally, there could also be a risk of market failure. Recognising property rights in personal data could not only encourage more trade in personal data and thus result in less privacy (Cohen 2010), it could lead to underinvestment in technology and services that enable the expression of privacy preferences. This would then result in greater information asymmetries between firms and individuals whose data is collected (Langer 2003; Schwartz 2004). Scholars have concluded that there is just no simple rule on whether privacy of personal data raises or reduces welfare as it depends on the circumstances (Hermalin and Katz 2006; Taylor 2004). However, it is commonly acknowledged that a free market in personal data will not provide an economically-efficient outcome. The degree of negative externalities within and across markets will depend on circumstances, as will any increase or decrease in welfare (Hui and Png 2006).

Personal data is widely used to create personalised offers such as products, prices, diets, recommendations, insurance, that are tailored to the characteristics of particular persons. There is much literature on whether personalisation improves exchanges and market efficiency, drawing from work on asymmetric information (Akerlof 1970; Spence 1973; Stiglitz 1975) and product differentiation (Mussa and Rosen 1978; Katz 1984; Moorthy 1984). It is no surprise therefore that new ways to gather more personal information would proliferate and their resulting data would find a market. Current regulation now implicitly acknowledge that personal data is a commodity, tradable and subject to the laws of supply and demand (Godel, Litchfield, and Mantovani 2012).

Mechanics of Personal Data Exchange on the Internet

The earliest form of digital personal data was derived from supermarket transactions, surveys and polls. These created the early data brokers that trade with firms, consumer data obtained from various sources, and individuals are generally not active agents in these transactions. Such transactions have had a market since the advent of ICT, one in which market research companies thrive, giving insights to firms based on their analysis.

The development of the Internet and the proliferation of e-commerce have resulted in an explosion of personal data supply and with it, public concern about privacy, as reviewed earlier. Personal data can be gathered from visits to websites, then used to analyse browsing and shopping behaviours. With cookies, data can be collected across all website visits and individuals can be easily tracked as they leave behind a data trail. With clickstream and identifying information, websites can profile visitors to a high level of accuracy.

Firms giving personal data to third party analytics services to obtain insights on their customers also contribute to the creation of more personal data. This is often covered by the firms' service terms of conditions to which individuals must agree. Insights are sometimes shared with customers themselves e.g. sleep quality by Fitbit.

Firms selling on personal data to data brokers also contribute to the market. This may be legal if the data has been suitably anonymised or even without anonymisation if the firm is based in a jurisdiction that does not have data protection laws. Finally, a 'market for privacy' exchange services where consumers buy technology to protect their personal information. In this market, some business models centre on bridging the market for privacy and the market for personal data by providing technology or services to give consumers more control over their personal data and also to provide them with opportunities to create more explicit exchanges of their personal data.

It is important to note that while a thriving market exists for personal data, a vast amount of personal data is not shared by firms. Employee data, students exam results, interactions on many smartphone apps are some of the data that have stayed within firms and have not been shared with third parties. However, with advancement in technologies and increasing API access, there is a growing sentiment that personal data liberated from these walled siloes could create greater innovation and opportunities. On the other hand, there is also increasing fear that liberating personal data would generate externalities that are socially inefficient, compromising privacy without any mechanism for internalisation.

The Value of Personal Data

Not all personal data is created equal and its availability in the market place depends considerably on its value. Generally, economists consider personal data or personal information, whether or not it is in the form of bitstrings, to be valuable in four ways.

First, personal data is used by firms to discriminate between segments with differing willingness to pay or reservation prices or for general segmentation of customers with different personal attributes (Mussa and Rosen 1978; Katz 1984; Moorthy 1984; Hart and Tirole 1988; Tirole 1988; Png 1998). Economists have concluded that price discrimination leads to socially efficient outcomes, and exploitation of personal data could lead to ex post inefficiencies, over investment in information (Hirshleifer 1971) and consumers being priced out of the market when there is more personal data available (e.g. Hart and Tirole 1988; Thisse and Vives 1988; Fudenberg and Villas-Boas 2006).

The second stream of economics literature on the value of personal information concerns the issue of a firm collecting personal data in one market for use by itself or selling it on to others in another market. In such cases, the firm would have big incentives to collect personal data, even at the expense of some of its own potential consumers (Taylor 2004). Taylor (2004) found that the option of selling personal data for extra revenue reduces social efficiency on the demand side in terms of loss in trades and increase in deadweight losses as well as on the supply side in terms of the cost (and risks) of collecting the data. Selling personal data into a different market also creates cross market externalities (Hui and Png 2006).

The third form of value of personal data comes from its use by firms to directly promote to consumers, often through unsolicited promotions such as direct selling, mail, catalogues, emails and advertising. Economists have considered such promotions as imposing costs of intrusion on individuals and view them as a direct externality (Camp and Osorio 2003), unrelated to the terms of a transaction or trading relationship (Laudon 1996). This raises the need, and in turn creating a market for seclusion, the need for solitude. Finally, the use of personal data also spurs the market for autonomy, the need to be in control of, or opt out from, surveillance and observation. Hui and Png (2006) argued for an understanding of all three aspects of privacy - secrecy, seclusion and autonomy - that personal data feed into. The fourth form of value, one that is the focus of this paper, comes from personal data use in real-time, on demand and dynamically connected markets.

Personal Data in Real-time, On Demand and Dynamically-Connected Markets

Much of the work on personal data goes back to a time when information flows were slower. With the recent Facebook/Cambridge Analytica incident, it is clear that personal data has moved on. The Internet has done much to expedite this, even into the physical space. Connected things have made available person-

generated data from the consumption of cyber or cyber-physical goods and services. This brings the physical world into the digital world at a phenomenal pace. The digital world now almost mirrors the physical world in terms of a way of life. Where in the past I woke up, went to work, shopped at malls and relaxed through physical leisure activities, the digital world now consumes a large part of human life in terms of work (many people work online), shopping (at Internet shopping sites) and relaxation (by watching videos and reading the news online). The Internet is no longer merely a 'channel' for the purchase of goods or information, but a medium of living that interacts with the physical (cf. Poster 2001). Data generated now include Expressions (e.g. words from social posts and messages), Spending (data from banking transactions), Photos (taken from the camera on the phone), Environmental (temperature and air quality from a wearable), Calendar events (from inputs into a digital calendar), Interactions (such as data from using a fridge or IoT device, or from listening to music), Activities (sport activity or sleep data collected from wearables such as Fitbit), Location (from a smartphone GPS) and Search (word typed into search engines), which I assign the acronym of ESPECIALS. The ESPECIALS set, which is by no means exhaustive, is an exemplar set where individual behaviours are 'visible' through data. In these cases, individuals are directly involved in generating the personal data and sometimes may not even know that the data is generated (e.g. when opening a door equipped with a sensor). The 'exchange' is part of a consumption activity and not an explicit exchange of data by the individual. The data is sometimes content data (e.g. words typed into a message) and sometimes metadata (the time the message was sent). This data has been growing in terms of value to data brokers, who can use it to profile a person's interest and daily living much more comprehensively as well as uniquely.

The availability of API access has made possible the sharing of data such as those in the ESPECIALS set with third parties, and personal data that used to be static and useful for research and insight is now available in real time, on-demand and dynamically updated. Thus, the ability to infer context, interests, preferences, priorities at every moment of a person's daily activity is now possible, if the data can be made available. API access is expanding, and more personal data collected and generated means a higher likelihood of increasing its use. Having real time, on demand and dynamic understanding of a person now fuels hyper-targeting, the ability to communicate directly with individuals in real time, on demand with various artefacts (e.g. ads, newsletters, articles, videos) across channels (social media, browsers) with dynamic messages and content. Hyper-targeting requires an understanding of the persona, which is aided by a copious amount of personal data. Hyper-targeting is essentially a product's choice of spatial locations (Hotelling 1929; Salop 1979) when the marginal cost of product is zero, which is the case for many cyber-social products such as news or music. These products are able to personalise algorithmically based on personal data updates.

This is now moving into physical products as well. As physical products embed sensors or software interfaces, the marginal cost of product *variety* becomes zero. Wathieu (2004) investigated product variety but only in terms of how enabling

privacy may increase the firm's advertising costs since the firm may not be able to segment demand using personal data. With a software/cyber layer, product variety can be achieved dynamically and on-demand (Yoo, Henfridsson, and Lyytinen 2010). Economists generally believe that producers should and will produce more and more of the good, up until market saturation, and competitive markets would drive the price of a good to zero. However, when the marginal cost of product variety is zero with personal data being infinitely expansible, the product can achieve perfect price discrimination and perfect spatial monopoly. While society may see too little innovation in Arrow-Debreu equilibrium, we are now seeing innovation in real time, aided by machine learning. In essence, this is what personal data fuels. To the extent that the demand for personal data is so high, firms would attempt to get more of it generation from the source. Personal data is now generated when a person explicitly gives it to firms such as through doing IQ quizzes or Personality tests in the case of Cambridge Analytica, or by asking cute questions to determine the person's context. I now move towards more formally articulating and theorising the economic properties of personal data as a digital good.

Theorising the Economic Properties of Personal Data

I consider a broader definition of personal data for the purpose of understanding its economics and define it axiomatically as the data *about* a person, and/or the data *generated by* a person. The former may be assessment reviews on an employee or CCTV camera footage of a person buying groceries. The latter may be data generated when a person opens a fridge door or reads an online newspaper. Each dataset has a source e.g. Facebook, Calendar or a door. Personal data, as bitstring data, shares some of the properties of normal digital goods. It is *non-rivalrous* in the sense that the consumption by the firm's customer does not prevent the firm itself from consuming it (Shapiro and Varian 1998). Second, bitstring data is *infinitely expansible*. That means a firm's bitstring data of a person can be copied to another space with very low marginal cost of re-production. Finally, bitstring data has a *non-excludability* property, implying that it is near impossible to exclude others from consuming the data. Non-excludability, however, depends on the legal and technological framework and it is entirely possible to exclude others from consuming through intellectual property rights, tied to the original copyright holder, or through a technical barrier such as encryption. If however, the original copyright holder could be persuaded to share, either by regulation or by consent of the person, it is entirely possible the market could be flooded with the same bitstring data. Formally, we can state this as $x_i(t) \in X^n = \{x_1(t), x_2(t), \dots, x_n(t)\}, \forall i \in Z^n, t \in R$ where $x_i(t)$ denotes the value of attribute i generated at time t within the dataset X . An example would be a post on Facebook where attribute $i = \text{post}$, $t = \text{time of post}$ and $x_i(t) = \text{content}$ (e.g. "I love wine!"). Note that in some services, the person generating the data may be unknown e.g. someone searching in an Internet cafe. If the person is known, for example the data was generated when the user was logged into an account, we can denote the data as $\{x_i^p(t)\} \subset X^n$.

Personal data as a digital good is also *recombinant* in that other digital goods can be created from combining them with features that are absent from the original, parent digital goods (Quah 2003). That means it is possible to create a new dataset A where $A = \{x_i^p(t), y_i^p(t), \dots\}$. An example is a set of Facebook posts in various location over time.

There is, however, one property that separates normal digital goods and personal data. Digital goods normally exhibit indivisibility or discreteness. This means that the consumption of a digital good such as music is usually the consumption of a whole good (the song); a fractional copy is worth very little. Personal data is divisible. For example, for a location dataset, the set of location data between 7am and 9am may be meaningful for firms who wish to understand commuting journeys. Furthermore, where a digital good like music is expensive to produce and cheap to reproduce, personal data is cheap to produce and cheap to reproduce, although excludability through the legal framework to protect consumers is increasing its costs. Finally, personal data can be discrete and bounded, such as a data point on age, or continuous, such as location data.

The property of divisibility together with recombining results in the ability of personal data to evolve a derivative which I term as *inferential signal*. In contract theory, signalling is a concept where one party conveys some information about themselves to another party. The most typical example is Spence's (1973) seminal paper on how acquiring formal education such as degrees is a signal of individuals' abilities to potential employers. The value of the degree as a signal comes from the fact that the employer believes the signal i.e. the degree is able to separate higher-ability from lower-ability employees, and therefore the employer is willing offer higher wages to degree holders, even if they do not know that to be true of a particular employee beforehand.

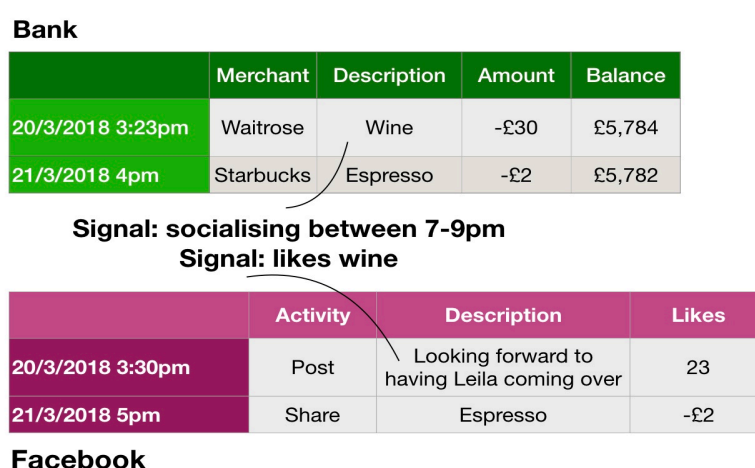


Figure 1: Signals are inferred from personal data

Similarly, personal data can form combinations where an inferential signal can be generated. An inferential signal could be *context* (in-town-shopping), *interests* (loves-wine), *priorities* (needs-more-sleep), *preferences* (wears-only-jeans), *decisions* (buy-shampoo), *intent* (going-to-Boston), *behaviours* (ate-Thai-food) or *persona* (savvy-digital-native). Thus, personal data divisibility, recombining and expansionability can result in the creation of infinite inferential signals.

Signal composition and quality. Signal composition is the composition of data within the signal. Within each signal the frequency of data, as well as the inclusion of other data could have a direct impact on the signal quality. For example, repeated orders of wine from supermarket purchases could be a signal that the individual likes wine; a Facebook posting of “The 2009 Medoc was excellent” is also a contribution to the signal. Thus the composition of each signal could also have the property of near-decomposability (Simon 2002), where intra-signal data frequency linkages may be stronger than inter-signal frequency linkages. Formally, we can consider an inferential signal of S defined as the image of a function from a personal data set A i.e. $f: A \rightarrow S$ and $A = \{x_i^p(t_j), y_i^p(t_j) \dots\}$

Simple Application of the model on Identity as a signal. Identification can be considered an inferential signal that is not time dependent. For example, it is commonly said that 87% of Americans can be identified through their birthdate, five digit zip code and gender. Thus, an identity inferential signal S for person α could consist of: $S_1^p = f(x_1, y_1, z_1) | x_1 \in X^n, y_1 \in Y^n, z_1 \in Z^n$ where $x_1 = \text{birthdate}, y_1 = \text{zipcode}, z_1 = \text{gender}$

However, identity signal composition can also be obtained from an individual's location between 11pm and 6am, and homeowner's data: $S_2^p = f(a_1, b_1) | a_1 \in A^n, b_1 \in B^n$ where $a_1 = \text{location between 11pm and 6am}, b_1 = \text{homeowner's name}$. Finally, identity signal composition can consist of only one data point $S_3^p = \{p_1\} | p_1 \in P^n$ where $p_1 = \text{passportname}$

This means signal can be obtained from multiple data sources. For each of the identity signals above, one could posit the quality, or the likelihood of the signal to be true, as: $0 < \rho(S_1^p) < \rho(S_2^p) < \rho(S_3^p) \leq 1$

This application is therefore the case of how data protection regulation (DPR) defines personal data. DPR definition of personal data specifies the data that relate to an identified or identifiable person or persons, a definition that is widespread across literature for the purpose of regulating its use. This application of our model therefore considers the DPR definition as personal data with an inferential signal of “personal identity”. It also raises an interesting issue of when personal data usage falls under the purview of enforceable regulation, perhaps when the probability $0.5 < \rho(S^p) < 1$.

Signal access and perishability. Conceptually, there are two types of signals that are valuable for the market. First, durable signals that are less time or context

dependent. These signals are, for example, identity, personality, marital status, demographic attributes etc. They are valuable because studies have found correlations between these signals and their propensity to buy certain goods and they are used to target customers. The second type of signal I define here is perishable signals. These are context and often time dependent, for example 'would like a cup of tea' or 'have 15 minutes before train leaves'. Such signals will perish either because the need is gone or it has been satiated. For the purpose of this paper, I denote all signals $S(t)$ as dynamic, time dependent signals $t_0 < t < t_1$ and t_0 is the time the signal is active and t_1 is when the signal is perished i.e. $S(t_0) = S(t_1) = \text{undefined}$.

Matching Inferential Signals: The Model

Perishable signals are precious commodity to firms, individuals and society. Combining personal data, machines and algorithms can improve human decision, action and societal coordination, especially when the action can be performed in real time and on demand. Combinations of personal data and algorithms can generate signals from early detection of depression from an individual's social media posting, signals from early detection of cancer or other diseases resulting from combining data from real behaviours and healthcare, or signals from using an individual's history to consider preferences and options for the future. Signals have gains for firms as well. First, since personalisation from infinite product varieties is possible with a digital layer at low marginal costs, there is much surplus to be gained by firms to position their offering to match the signals if they are precise, good quality and the match occurs before perishing. Second, consumer need for a product (e.g. insulin or tea) may not be known even to the individual until the need or urgency arises. Perishable signals allow markets to form outside of conventional markets and enable exchanges to occur, surfacing latent need and increasing willingness to pay (Ng 2014).

Formally, we construct a matching game as follows:

Consider a person being asked to share data by signing on to an app. The person can choose to share or not share. At time t when a signal is active, the firm decides to offer a service, a discount or perhaps a message of advertising. Nature then decides on whether the signal is matched, in which case the person obtains a payoff of $\pi_p - v$ where v is the vulnerability cost that comes with sharing data (privacy, seclusion etc.) and the firm obtains a payoff of $\pi_f - c$ where c is the cost of matching.

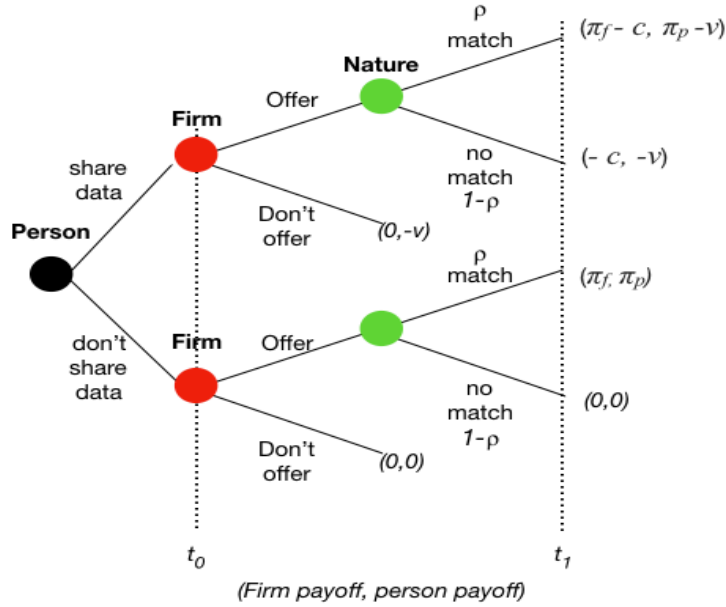


Figure 2: Formal game of perishable signal matching

If the person decides not to share, the firm can also decide to offer or not offer the service and the payoffs are the same except without vulnerability cost, since data is not shared, nor the cost of matching.

We can depict the payoffs in a matrix set out below.

| | | Firm | |
|--------|-------------------|--|--------------|
| | | Offer | Do not offer |
| Person | Share data | $(\rho\pi_f - (1 - \rho)c, \rho\pi_p - (1 - \rho)v)$ | $(0, -v)$ |
| | Do not share data | $(\rho\pi_f, \rho\pi_p)$ | $(0, 0)$ |

Proposition 1: The firm will always offer and the person will always prefer not to share data

The above payoffs show that because $\rho > 0$, the person knows that the firm will always offer. Since the firm will always offer, the person would prefer not to share data and the equilibrium strategy will be (do not share, offer).

From the above proposition, the firm would certainly change its strategy. This is usually the case for many apps that force users to sign up before they can even begin to use the service. In so doing, the Nash equilibrium is can be designed to become (share, offer).

| | | Firm | |
|--------|-------------------|--|--------------|
| | | Offer | Do not offer |
| Person | Share data | $(\rho\pi_f - (1 - \rho)c, \rho\pi_p - (1 - \rho)v)$ | $(0, -v)$ |
| | Do not share data | $(0, 0)$ | $(0, 0)$ |

The problem when this happens is that when signal quality drops very low, the firm will not offer.

Theorem 1: When $\rho \rightarrow 0$ Perishable signals face market failure

The logic is simple, when $\rho \rightarrow 0$, The firm would prefer not to offer so as not to incur the cost c . The person will always prefer to share data because he will (weakly) prefer to have a positive payoff to none at all.

This simple model shows that it doesn't not require privacy concerns for markets to fail. Even when there is willingness to share data and willingness to create signals, market failure arises due to inability to match before a signal perishes.

As more personal data become available, in real time and on demand, signals have become more accurate, fuelling the advertising, geo targeting and location-based marketing economy. In a digitally-connected society, real-time signals perish quickly, which in turn fuels more intrusive data collection that would keep signals proliferating. As regulators react to the Facebook/Cambridge Analytica scandal, access to real-time, dynamic personal data and data signals is becoming a challenge as economic, regulatory and privacy concerns will drive up costs for personal data exchanges, including social costs. As a resource, personal data that is valuable for real time, on demand and dynamic signalling runs the risk of being more scarce, less real time, less on demand and less dynamic. This not only means that the incumbents that hold it become more powerful, opportunities of un-exploitable data opportunities for the good of individuals and society, lost. As technology

advances, the infrastructure that creates this value is showing its weaknesses and vulnerability. In the face of an increasingly-empowered population of connected citizens, the centralised organisation control of personal data has become restricting.

Work in market design through “microeconomic engineering” (Roth 1991) have shown that transactions and institutions matter and could be redesigned to engender better market outcomes. Given the market failure, a micro-economic redesign of the personal data exchange market may be a solution. The next section reports the micro economic redesign and engineering of the personal data exchange and its implementation.

Legal and Technological Design and Creation of the HAT (Hub-of-all-Things) for Creating Person-controlled Personal Data (PPD) Asset Class

Property rights is said to be the most important factor for markets of a good to exist. This is because markets not only enable the exchange of a good, but trade the various exclusive rights associated with the good in terms of its use, exclusion and alienability (Demsetz 1967; Alston, Libecap, and Schneider 1995, Carruthers and Babb 2000). For the exchange of a digital good such as personal data to occur in a meaningful way, some exclusive rights must exist with the owner to exclude others from making arbitrary copies of the data. This is the case also with personal data controlled by firms, which I term OPD (organisation-controlled personal data). OPD access rights are granted to other organisations through API access terms and conditions or through permitted selling of data to data brokers. I define PPD as *person-controlled personal data*, the personal data where intellectual property rights and excludability of the data (control) is with individuals. Since digital personal data consists of bitstrings, and is created by the technology that collects it, it is possible that rights could be retained by individuals if they legally owned *a technological artefact capable of real-time, on-demand and dynamic exchange of personal data*. In addition, since personal data is non-rivalrous and infinitely expandable, copies of the data may be accorded with a different set of *legal rights*, rights that could be controlled by individuals themselves. Finally, since signals are generated from combinations of personal data, the quality of the signals would depend on the ease of different personal datasets to be combined, bundled and exchanged so that the *economic value of the data signals* are high.

To make the PPD asset class a reality, I present the design and built of the HAT (Hub-of-all-Things) technology and legal artefact.

Technology. For brevity, the HAT technological architecture can be seen at https://developers.hubofallthings.com/home/tech-stack/HAT_core.html. In essence, the HAT is designed and constructed as private, standalone databases embedded within containerised microservices for personal data that have clear boundaries of data at rest, data in transit and data in use. Aside from just being a

datastore, containerised microservices wrapped around the database means it can be a “micro-server”, capable of processing data within. By isolating each HAT micro-server from one another so that every HAT is one containerised microservices-enabled database per person, boundaries are clear and rights can be bestowed. Technologically, the schema (data structure) was chosen to be flexible for outbound data, but keeping the rigidity of inbound data. Apps that give data into the HAT retain their original table and data structures within their *namespaces* e.g. Facebook namespace in the HAT has a Facebook table of data, same with their original names, similar to Spotify, Google Calendar, Fitbit etc. Outbound data from the HAT, however, can allow infinite combinations of data values across datasets e.g. Twitter tweet at 3pm with iPhone location and heartbeat. Each of these data values and bundles can be named and then exchanged through standard APIs using standard Internet protocols and encryption in real time. Within the HAT comes the technological capability of embedding functions e.g. pre-trained machine learning algorithms that transform data within the HAT and generate new data signals that sit within the HAT, which can be exchanged through the standard APIs if the individual wishes to.

Legal Rights. Personal data use contracts cannot specify all states of nature or all future actions and use of the data in advance. When there are states or actions that cannot be verified ex post by third parties, they are therefore not possible to be contractible ex ante. This means that the contract must include discretion and that discretion is to be exercised by whoever is allocated the ‘ownership’ rights to the personal data. The literature on incomplete contracts (see Grossman and Hart 1986; Hart and Moore 1990; Aghion and Bolton 1992; Dewatripont and Tirole 1994) has typically focused on the question of which party in a contract should have the right to undertake certain actions in the management of those assets. If contracts were complete, it would make no difference who was allocated that right, since actions taken are a function of the situations stipulated in the contract. However, incompleteness of personal data contracts matter in terms of who has the power to take action, and the presumption is always that the economic actors will do so according to their interests. Deciding who should have the power to take certain actions is therefore a matter of foreseeing which actors will be most likely to act in the desired way. The allocation of power matters when it is not possible to specify in advance precisely how that power should be exercised. In the case of personal data, it is clear that future usage of one’s personal data must rest in the control of individuals, because this would reduce the incentive for the firm to sell on the data, especially when such an action may be obscured from the individual.

Legal ownership of rights to the micro-server can be bestowed due to the presence of clear boundaries resulting from HAT containerisation (see appendix A). Owners of such an artefact can therefore be afforded all of the intellectual property rights of the micro-server, reducing ambiguity of personal data use. Containerising one individual’s data within his own database wrapped with microservices allows individuals themselves to be a ‘data controller’ and ‘data processor’. Finally, the HAT core technology is uploaded to GitHub under an open-sourced AGPL license

(not be closed even if built upon), ensuring that any code within the HAT, which reveals how data is being handled within the HAT, is transparent to all.

In terms of usage of data within with HAT, it is clearly necessary for the data to be unencumbered so that full excludability rights are retained by the individual through the HAT and not by the source of the data. In this manner, micro-server access to data from the current sources such as Google or Facebook could be considered as subject access, whether directly or through a third party, which, under European law, suggests that the ensuing data retrieved by individuals are owned by them. While this is legally not proven in case law, a case can be made that micro-server owners have rights to reuse and reshare their own data within it as co-producers of the same data. Given that data within the boundaries of the firm cannot be meaningfully “propertised” by individuals, a non-rivalrous copy of the same data within a HAT micro-server with the same intellectual property rights on the copy for individuals as the rights of the source for the firm, can be an equitable arrangement.

I now set up a series of propositions for PPD. Instead of proving the propositions formally, the propositions are taken into the design of an artefact and the redesign of the market for personal data which I would elaborate on.

Datasets. As presented earlier, signals require access to personal data. In the current state of the personal data economy, datasets are contained within the siloes of ‘apps’. Each app could hold simple data of booking travel (Trainline) or choosing the clothing to wear (Stylebook), or managing a diet (MyFitnessPal). The combined data from all the apps could create much better signals. However, only some apps sell on the data and the regulatory climate of selling data is creating higher costs and risks. This suggests that the personal data market is narrowing and privacy concerns may hand greater monopolistic power to incumbents that hold much more data. The risky environment results in hoarding, since access is a challenge, creating greater security risks. However, if a micro-server could deliver outsourced benefits for giving individuals their own data such that it can still be accessed by firms in real time and on demand, cross-demand for data would fuel more datasets coming into the HAT micro-server.

Proposition 2: PPD can generate more datasets i.e. $Nargs\{S_o\} < Nargs\{S_p\}$

Signal quality. Since transactions of personal data and their signals do not involve the person, externalities occur. This is not merely with regards to privacy but also data quality, since the person is not a stakeholder of the data being transacted. With PPD, matched signals are actioned upon by individuals themselves, hence they are incentivised to ensure that the data is accurate. In addition, data brought into a micro-server is liberated from its structures, such that different combinations of data from the different structures could be created, resulting in better signals.

Proposition 3: Ceteris paribus, PPD creates higher signal quality i.e. $0 < \rho(S_o) < \rho(S_p) < 1$

Signal buyers. Signals are expensive, since purchasing, consolidating, aggregating and analysing personal data for sale is costly. With greater regulatory controls there is greater scarcity and the price for personal data (and their derivative insights or signals) is increasing. Hence, smaller apps that could better tailor or personalise their products through signals could be priced out of the market. However if signals belong to individuals, being able to reuse and reshare signals for matches results in lower marginal costs for firms.

Proposition 4: PPD creates pareto efficient outcomes

Signal matching and perishability. “Matching” is the focus of economic models on who does what and who gets what, particularly when a good is scarce and allocation is an issue (Niederle, Roth, and Sönmez 2008). For OPD, data brokers buy data and generate insights often long after the data is generated. This means only some signals are available, those which are stable and less time reliant. However, a HAT micro-server that is real time and on demand could obtain matches before it perishes, driving the demand for more signals and better signal matches.

Proposition 5: PPD Matching is more stable and efficient compared to OPD

Source constraints. While personal data could be accessible through open APIs by third party services sharing amongst one another, data collectors impose constraints on its use, creating legal excludability. This restricts what personal data can be used for. With PPD from HATs, individuals do not face such constraints.

Proposition 6: PPD reduces source constraints on excludability.

Given the above propositions, we propose that transaction costs for PPD would be low and a bargaining solution could exist (Coase 1960). What is then needed is to establish the market for PPD. We now turn to the economic design of the HAT data exchange system.

Microeconomic Design and Engineering of the HAT Data Exchange System for the PPD Market

Potentially, PPD could compete with OPD, at least at the beginning. Over time, PPD could prevail due to the quality of the asset class (as per propositions above). The current personal data market is imperfect, asymmetric and incomplete. Imperfect because of the ability of firms to create monopolistic competition with an extensive variety of cyber-physical products that can be personalised with personal data. Asymmetric because firms often have extensive information of consumers online through cookies and third-party data to price discriminate and target, and incomplete because follow-on uses of personal data are uncertain and ambiguous. PPD could make personal data markets more efficient by reducing asymmetry and

the ability to create complete contracts due to its accessibility in real time and on demand.

PPD Supply. For a well-functioning market, PPD supply must be in place; this supply of data into the HAT, thus becoming PPD, must then have the capability to be stored, processed, transformed into data signals and exchanged in real time and on demand unencumbered by the source. Today, Facebook, Google, Fitbit and many Internet services allow API access of data by third-party services primarily to create lock-in, engagement and greater network effects from combined services sharing of personal data. Access by an individual is therefore technologically possible, although legally ambiguous. In addition, European General Data Protection Regulation (GDPR) coming into effect on 25 May 2018 compel the need for data portability in individuals' right of access.

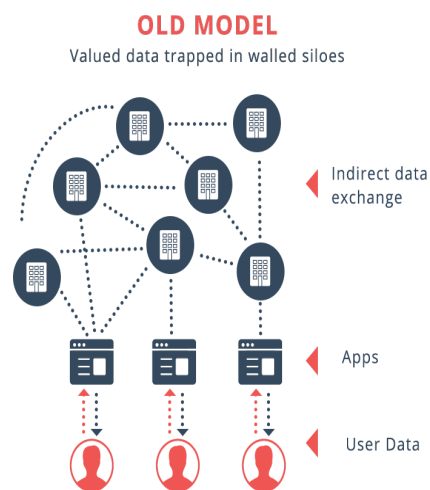


Figure 3: Value trapped in walled silos

While the law is vague on whether this means API access (thereby giving individuals real-time, on-demand and dynamic access), the fact that firms share personal data between themselves with consent would suggest that mounting a legal challenge for individuals' right to API access may be winnable, if legislators do not compel firms to do this. Notwithstanding the law, signals that are most valuable e.g. intent, preferences and priorities for example, can already be created with only ESPECIALS data, most of it available through API access already. Granting API access to an individual's HAT is merely an indirect mechanism to companies wanting a lock-in and a network effect, but inserting a new right of excludability (control) exercised by the individual.

PPD Demand. The creation of the PPD asset brings about the challenge of designing coordination between economic actors for more efficient and socially-optimal outcomes. Coordination has been said to be the economic problem that needs to be addressed (Knight 1951, p6; Leijonhufvud 1981, p321-322). In the way that the PPD asset is analogous to digital labour, a market-based coordination structure with interactions of demand (for PPD and PPD signals) and supply (of accessible personal data) could be a reinforcing loop, generating matches that can cause a spontaneous and ongoing coordination of separate economic activities between individuals and firms. For that to happen, there must be *thickness* (Niederle, Roth, and Sönmez 2008), a condition where a sufficient number of participants come together to transact. In the current market, data brokers have amassed a sizeable number of buyers interested in personal data and insights². The proliferation of the ‘app economy’ has created marketplaces that guarantee personal data supply for OPD and the \$61billion advertising economy has created a demand for OPD that could help influence, target, promote and sell to the millions who are online. PPD would need to compete in that market. The PPD market has to also overcome *congestion* by making transactions fast enough and yet having enough alternatives to choose from. For OPD, this is currently done through API access, and a distributed network on the Internet, often out of the jurisdictions of territorial governments. Finally, the PPD market must be *safe* to transact which means it is more optimal to transact within the marketplace than outside (Vulkan, Roth, and Neeman 2013).

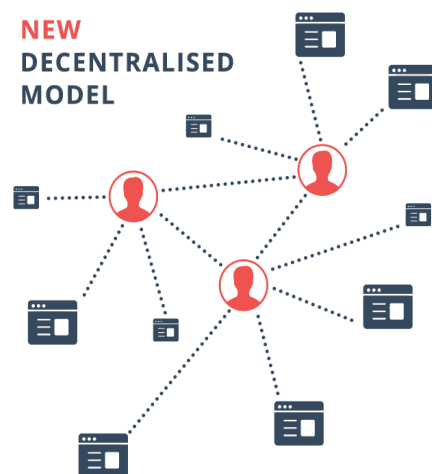


Figure 4: Decentralised model to unlock better signals

PPD Market Strategy

There are currently more than a million apps on devices such as the iPhone and Android phones. Every app downloaded and used by an individual has a direct

² <https://www.ftc.gov/reports/data-brokers-call-transparency-accountability-report-federal-trade-commission-may-2014>

relationship between the firm and the user. Each dataset of app personal data sits within each app, walled and silo-ed. Many of such apps belong to companies that are small, only a few large ones are the likes of Facebook and Google. All apps could benefit from more personal data and in particular, personal data signals. If a signal is about wanting coffee, a coffee app already on a phone may be interested to acquire it. If it is a need to buy an item, or medicine, apps within a person's phone are already spatially (location and time) within the reach of individuals.

This means that signals can be best matched by the firms of these phone apps. These apps have an in-app market of their own - between a firm and its users. In addition, these apps generate their own personal data whether it is train booking or wardrobe assistance. If the HAT is able to take on namespaces for the app data and also transform it privately to create new data signals to be bought and consumed by the app, thickness, congestion and safety can be achieved within every app as an internal market. Thickness, because the app has an existing user base to sell to. Congestion, because signal transactions can follow the same manner of transacting data within the app and safety, because the firm has already a relationship with the individual through the app. Such an internal data exchange can be a marketplace setting for PPD and PPD signals.

Proposition 6: Creating exchanges of PPD within a firm's internal market of its own customers can (1) create thickness; (2) reduce congestion; and (3) create an environment of safety

What this implies is that PPD and PPD signals are brokered for sale between a firm and its own customers within the firm's own app, rather than creating a different marketplace. This could be implemented through the firm issuing 'private data accounts' to its own customers with the intent of buying PPD or PPD signals. As other apps begin to store data within the person's micro-server, firms can request data not generated by them, thus enabling the individual to be a multi-sided market for PPD.

Proposition 7: Decentralised PPD generates network effects through resharing and reusing PPD between apps and firms

Market Redesign: Implementation and Early Findings

The above propositions were implemented technologically, in tandem with the microeconomic model; the HAT was created as a one-database per person wrapped with containerised microservices and accorded the HAT owner with full legal rights to the database and the data within (thus creating PPD). Services built to host HATs on the cloud and the exchange of PPD data between firm and individuals were built as API services, fully encrypted end to end, preserving the privacy of the HAT owner (even from the host) and giving full control to individuals. The HAT platform was rolled out through the formation of a company, HAT Data Exchange (HATDeX). HATDeX built service for HAT owners to view, use, control and

exchange data easily with toggles and button presses. Individuals pulled their data into the HAT through “Data Plugs” and shared through “Data Debits” analogous to the way individuals use their bank accounts, as a scaffolding strategy to assist in the understanding of the HAT. Coordinating mechanisms were implemented by communicating to firms how firms themselves could give their customers HATs by offering their existing customers base a ‘private data account’, complete with the firm’s labelling and branding, “powered by the HAT”, thus enabling decentralisation of personal data storage and exchange without third party services and the firm could retain ownership of the relationship with their own customers, while being able to obtain data signals. Following the prescribed microeconomic design, HATDeX created data plug services to generate non-rivalrous copies of OPD in the HAT, therefore enabling individuals to claim their own data from other sources on the Internet in real time. As of 20 April, 2018, HAT owners have PPD of real-time data from Spotify music listens, Google Calendar, Facebook and Twitter posts and Fitbit activities. Operationalising the design further, the control and processing of personal data are done within the HAT, enabling HAT owners to act as “data controllers” and processors of the data in their HATs. In the implementation, a decision was also taken to reduce the risks of lock-in by releasing the HAT open source code to ensure the transparency of data handling within HATs. HAT providers do not have any access to HAT data except through data debits for the data they wish to acquire. PPD Signals such as verified action (action verified by data coming into the HAT) was the first signals to appear and be shared.

The design of the legal, technological and economic system as set out above was implemented in January 2017 in beta and in full roll-out in November 2017. As of 20 April 2018, a total of 1104 HAT owners have signed up, with a 4.3% weekly growth. HATDeX reported 12 organisations negotiating deployments and 27 business leads in various stages of meetings. Firms that came forward were keen to leverage the value of personal data to improve their strategic advantage – they wanted to deliver engaging, high-value apps and services to their customers, but found it a challenge to do so.

Market Disruption

The disruptive effects of the implementation were recognised through reports in the *Financial Times*, *Wired Magazine*, *Spectator* and various news media³. The HAT has been mentioned at governmental levels, from an Australian government report on data collection and use, to the UK Lords Select committee on AI. The HAT’s disruptive potential was seen as the start of a digital consumer economy about to decentralise. Firms that came forward to adopt the technology did so for various reasons.

³ <https://www.hatcommunity.org/our-voice>

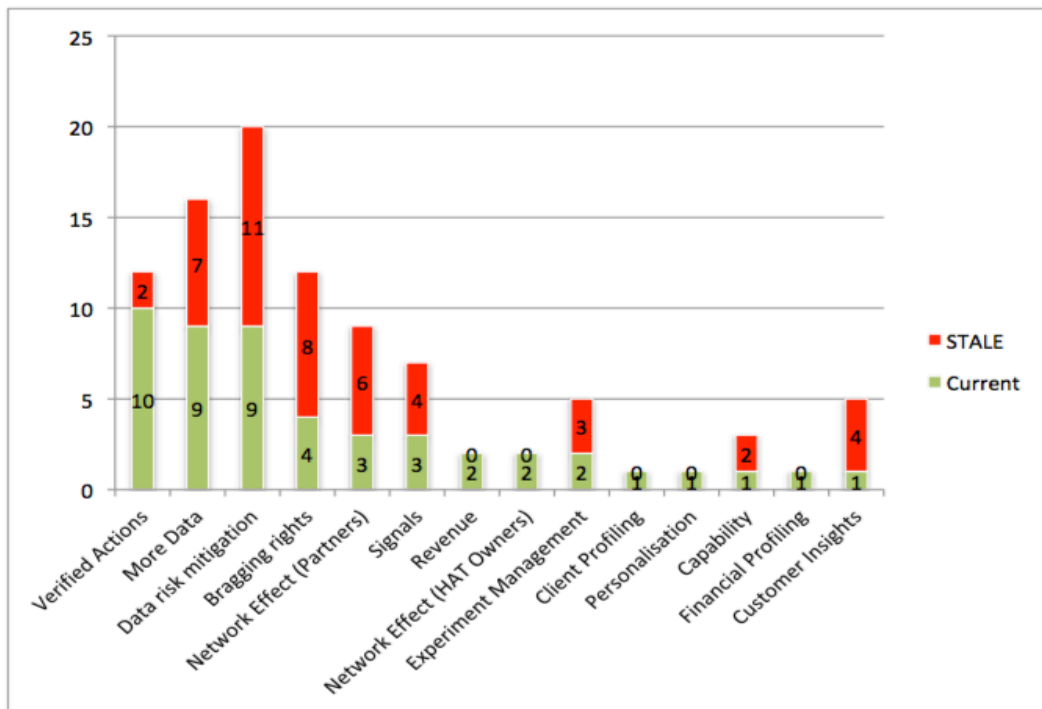


Figure 5: Firms' motivation for adopting HAT technol

The findings indicate their motivations (See Figure 5), of which there were 14 in total: (1) the HAT enabled firms to create a direct relationship with their own customers and their data so that firms can provide insights (customer insights); (2) being able to profile algorithmically without privacy risks was an advantage (financial profiling); (3) since holding personal data is costly, new apps benefit from not having to build user account management by building their services on the HAT, as they can be assured of data access on demand, as long as the data contract is valid (capability); (4) the ability to personalise their offering around personal data was valued (personalisation); (5) they sought to provide their own clients (B2B) the ability to profile end customers (customer profiling); (6) they sought to run market research and experiments with real behavioural data (experiment management); (7) being able to share user base with a family of apps (network effect, HAT owners); (8) being able to broker customers data for revenues (revenue); (9) the ability to install private analytics and machine learning algorithms to generate signals for recommendations; (10) benefiting from the partners on the platform (network effect, partners); (11) bragging rights to being consumer champions of privacy; (12) mitigating personal data privacy risks; (13) the opportunity to request for more data than they can otherwise collect (more data); and (14) the opportunity to authenticate and verify actions such as listening to certain types of music, buying from certain shops etc.

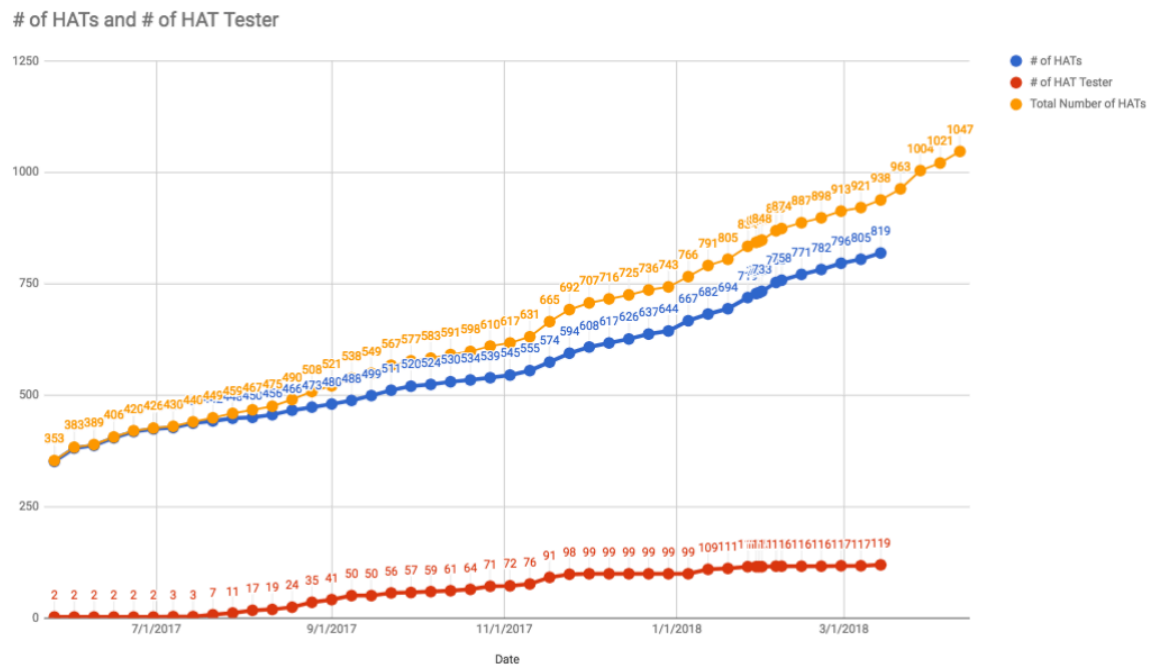


Figure 6: Growth of HATs before scaling

Consumers who downloaded retail HATs did so for four reasons: (1) control over their data; (2) being able to store and track their data; (3) wanting to monetise their own data; and (4) being evangelists for personal data empowerment.

In May 2018, HATDeX announced the creation of the HAT infrastructure Platform in collaboration with Tolga Uzuner, a global technology investor and former Partner at Apollo Global Management LLC with a £30 million planned commitment to create at least 10 million new HAT micro-servers, provisioning HAT micro-servers, growing technology infrastructure, and improving innovation in the decentralised data economy through 2020.

Discussion

While it is too early to judge if indeed the market for the new PPD asset is sustainable, research is ongoing in this space and the scaling-up of HATs in the coming months will provide empirical evidence on the type of PPD and PPD signals that would be in demand.

Proposition 2 is largely proven as HAT has now taken on multiple datasets partly from data plugs and partly from new data generated by HAT owners themselves through apps. Higher signal quality is evident through the ‘verified action’ signal

where HAT owners can accept offers to tweet, post or listen to music and the data generated from the digital action of actually tweeting and posting on social media allows the HAT to verify that the action was undertaken and send out the signal if the HAT owner accepts the signal exchange from an app. Pareto efficiency is yet unproven. As HAT scales to 10 million in the next two years, this may be measurable. PPD Matching while yet to be evidenced, and is partially validated by the companies in the process of integrating with HATs for the purpose of buying signals while mitigating privacy risks. Source excludability is legally untested, although unofficial conversations with lawyers and government suggest that HAT data is unencumbered. The findings showed that the choice of coordination structures clearly made a difference in attracting firms. That HATDeX technology was infrastructural and created integration mechanisms not merely at the technological, but also at the business and experience levels, showed that transactions designed were able to scaffold into the existing markets and evolve them into new ones. Seeing one app's data in another app, even if it is the person's HAT dashboard app, is the most mind-bending realisation for many firms and individuals. Reusing and resharing for PPD network effect value is seen to be most disruptive and the findings suggest that this polarised the firms. Some firms used HATs exclusively for their apps, not bothering to create user accounts at all and are willing to share with others, while others valued having personal data both on their apps and in HATs, suffering some discomfort of 'letting their data go' despite the law in Europe mandating that individuals have rights to access their own data. Having consumers with PPD accessible real time and on demand has created alarm bells for many firms, as the fear that consumer control means a loss of market advantage, since their data could be given to others. Other firms this viewed positively as they saw their app data being used by other apps as the potential to create lock-in effects. In terms of matching and creating a market for PPD, serving internal markets of firms and their existing customers helped to create a safe environment for transacting PPD, and using the firm's existing revenue models reduced congestion. It also made matching easier, as signals were aligned to what both firms and their customers wanted within one app. By targeting firms with existing customers, internal markets enabled PPD to achieve thickness with lower effort.

Economics of Personal Data. Personal data is a production asset to create greater product variety at low marginal costs. Potentially, this could be treated the way all production assets are treated - as resources for the production of goods such as land, labour, capital (Samuelson and Nordhaus 2004). In an age of Internet-Connected Objects (ICO) where data is generated from a doorbell ringing, water consumption from a shower, the Kindle page on which the reader has stopped reading; zetabytes of bitstring data is being generated with or without an individual's direct involvement, whether it is useful or not. The issue of whether bitstring data as a production asset should derive its income from rent (bitstring data coming from land) or interest (bitstring data from capital goods) or wages (bitstring data from human labour) becomes a relevant discussion. If typing into a search box is digital labour, then search results can be deemed as 'wages'. The current situation is that all bitstring data belongs to the capital asset that generates

it, regardless of the creator of the data. This means that the data generated by a tractor would belong to the manufacturer – rather than the owner of that tractor, since the current rules on bitstring data generation and rights do not differentiate between bitstring data as an asset, or the asset itself. The concepts, theories and fundamentals of personal data as an asset needs a robust economic discussion.

Conclusion

The technological, legal and economic architecture of a containerised private database with microservices - a personal micro-server - has a few advantages. Through the microservices, individuals can exchange the personal data within their database for their own benefit, deriving income from it or transferring it for fun or service if they wish. The containerised microservices help individuals do this, using standard APIs, with the individuals themselves staying in full control. The individual can become an effective on-demand data supplier to firms building services that require personal data. Personal data form entry, personalised quotations, assessments, online identity verification, and user account creation all carry risks and costs for the corporation. An alternative that allows for the sharing of personal data, by individuals themselves, from their own personal micro-servers can save both businesses and individuals time, effort, risk, expense and liability. Widely adopted, the number of personal micro-servers could greatly reduce the incentives for cyber-attacks. A penetration into one secure database container yields the perpetrator of that attack exactly one database, where in the current system a similar risk would yield up to billions of records of personal data instead.

This paper theorised the fundamental economic properties of personal data to argue the challenges and vulnerabilities faced by organisation-controlled personal data. It then set out a series of propositions for market redesign to enable a more efficient technological, legal and economic model through the creation of PPD, and argued for its usage especially in AI, health and wellbeing. With the HAT micro-server and PPD, privacy is no longer an issue. This paper proposes that economic risks would be lowered with PPD and personal data signals can proliferate, creating new services using personal data and signals in real time and on demand. Individuals can share their signals for 5 minutes or 50 days and with more access to data from the European General Data Protection Regulation, more datasets can come into individual HATs for privacy-preserving tools and algorithms to flourish, generating even more signals and better signal quality. The HAT also takes AI to the next level - *Aug-I - Augmented Intelligence* - enabling a combination of personal data, machines and algorithms to work together to improve human decision, action and societal coordination - all in real time, with low privacy risks. Since the data is transacted with the individual and not a third-party broker, and the signals generated can create recommendations and personalisations that are relevant, the person becomes a stakeholder in the quality of the data, and remains in control of what is shared. Since data is available in real time and on demand, with low marginal costs, we propose that more signal buyers can emerge from new apps created. And since HATs are always on, data scarcity is reduced and smaller players

with low data science capabilities can come into the market to create apps. This means also that hoarding of personal data can reduce, creating a more secure world. With always-on API access, signals can also be matched with recommendations and personalisations before they perish. Finally, since the HAT owner owns the micro-server and therefore the rights of the data within, there are no data source constraints and the HAT owner benefits from a historical account of his digital life that could be shared for better societal health and wellbeing. As the HAT platform scales, further research aims to obtain an empirical understanding of PPD exchanges and how the market evolves, especially for data signals. This would enable an understanding of the value of PPD to HAT owners, organisations and indeed, society, to the level of data granularity that has never been possible.

References

- Acquisti A (2010) The Economics of Personal Data and the Economics of Privacy. Background Paper for OECD Joint WPISP-WPIE Roundtable, 1.
- Acquisti A, Taylor CR, Wagman L (2016) The Economics of Privacy. *Journal of Economic Literature*, 52(2); Sloan Foundation Economics Research Paper No. 2580411.
- Aghion P, Bolton P (1992) An Incomplete Contracts Approach to Financial Contracting. *The Review of Economic Studies*, 59(3): 473-494.
- Alston LJ, Libecap GD, Schneider R (1995) Property Rights and the Preconditions for Markets: The Case of the Amazon Frontier. *Journal of Institutional and Theoretical Economics (JITE) / Zeitschrift Für Die Gesamte Staatswissenschaft* 151(1): 89-107.
- Akcura MT, Srinivasan K (2005) Research Note: Customer Intimacy and Cross-Selling Strategy. *Management Science*, 51(6): 1007-1012.
- Akerlof GA (1970) The Market for 'Lemons': Quality Uncertainty and the Market Mechanism. *Quarterly Journal of Economics*, 84(3): 488-500.
- Anderson SP, de Palma A (2005) A Theory of Information Overload. Unpublished manuscript, Department of Economics, University of Virginia.
<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.487.2261&rep=rep1&type=pdf>
- August T, Tunca T (2006) Network Software Security and User Incentives. *Management Science*, 52(11): 1703-1720.
- Belleflamme P (2016) The Economics of Digital Goods: A Progress Report. *Review of Economic Research on Copyright Issues*, 13(2): 1-24.
- Beresford A, Kübler D, Preibusch S (2010) Unwillingness to Pay for Privacy: A Field Experiment. IZA Discussion Paper No. 5017.
- Camp LJ, Osorio CA (2003) Privacy Enhancing Technologies for Internet Commerce. Petrovic O, Ksela M, Fallenböck M, Kittl C, eds. *Trust in the Network Economy*. (Springer-Verlag, Berlin), 317-332.
- Carruthers BG, Babb SL (2000) *Economy/society: Markets, Meanings, and the Social Structure*. (Pine Forge Press, Thousand Oaks, CA)
- Charter of Fundamental Rights of the European Union, OJ C 364, p. 10, 18.12.2000, Article 8

Coase RH (1960) The Problem of Social Cost. *The Journal of Law and Economics*, 3:1-44.

Cohen JE (2012) What Privacy is For. *Harvard Law Review*, 126(7): 1904-1933.

Commons JR (1931) Institutional Economics. *American Economic Review*, 21: 648–57

De Mauro A, Greco M, Grimaldi, M (2016) A Formal Definition of Big Data Based on its Essential Features. *Library Review*, 65(3): 122-135.

Demsetz H (1967) Toward a Theory of Property Rights. *American Economic Review*, 57(2): 347-59.

Dewatripont M, s Tirole J (1994) A Theory of Debt and Equity: Diversity of Securities and Manager-Shareholder Congruence. *The Quarterly Journal of Economics*, 109(4): 1027–1054.

Fudenberg D, Villas-Boas JM (2006) Behavior-Based Price Discrimination and Customer Recognition. Hendershott T, ed. *Handbook on Economics and Information Systems* (Elsevier, North-Holland, Amsterdam).

Grossman SJ, Hart OD (1986) The Costs and Benefits of Ownership: A Theory of Vertical and Lateral Integration. *Journal of Political Economy*, 94(4): 691-719.

Gandomi A, Haider M (2015) Beyond the Hype: Big Data Concepts, Methods, and Analytics. *International Journal of Information Management*, 35(2): 137-144.

Godel, M, Litchfield A, Mantovani I (2012) The Value of Personal Information: Evidence from Empirical Economic Studies. *Communications & Strategies*, 88(4th Quarter): 41-60.

Hann IH, Hui KL, Lai YL, Lee TSY, Png IPL (2006) Who Gets Spammed? *Communications of the ACM*, 49(10): 83-87.

Hart OD, Tirole J (1988) Contract Renegotiation and Coasian Dynamics. *Review of Economic Studies*, 55: 509–540.

Hart OD, Moore J (1990) Property Rights and the Nature of the Firm. *Journal of Political Economy*, 98(6): 1119-1158.

Hermalin B and Katz M (2006) Privacy, Property Rights & Efficiency: The Economics of Privacy as Secrecy. *Quantitative Marketing and Economics (QME)*, 2006, 4(3): 209-239.

Hirshleifer, J (1971) The Private and Social Value of Information and the Reward to Inventive Activity. *American Economic Review*, 61(4): 561-574.

Hotelling, H (1929) Stability in Competition, *Economic Journal*, 39(153): 41–57.

Hui, KL, Png IPL (2006) Economics of Privacy. Hendershott T, ed. *Handbook on Economics and Information Systems* (Elsevier, North-Holland, Amsterdam).

Katz ML (1984) Firm-specific Differentiation and Competition Among Multiproduct Firms. *The Journal of Business* 57(1): S149-166.

Knight FH (1951) The Role of Principles in Economics and Politics. *The American Economic Review* 41(1): 1-29.

Laudon KC (1996) Markets and Privacy. *Communications of the ACM*, 39 (9): 92-104.

Leijonhufvud A (1981) *Information and Coordination: Essays in Macroeconomic Theory*. (Oxford University Press, Oxford) 321-32.

Loebbecke C (2002) Digital Goods: An Economic Perspective. Bidgoli H, ed. *Encyclopedia of Information Systems* (Academic Press, San Diego) 635–647.

Langer R (2003) Where a Pill Won't Reach. *Scientific American*, 288(4): 50-57.

McGeveran W (2001) Programmed Privacy Promises: P3P and Web Privacy Law. *NYU Law Review*, 76(6): 1812-1854.

Moorthy KS (1984) Market Segmentation, Self-Selection, and Product Line Design. *Marketing Science*, 3(4): 288–307.

Mussa M, Rosen S (1978) Monopoly and Product Quality. *Journal of Economic Theory*, 18(2): 301-317.

Odlyzko A (2003) Privacy, Economics, and Price Discrimination on the Internet. *Proceedings of the 5th International Conference on Electronic Commerce* (Pittsburgh, Pennsylvania) Sept 3-Oct 3, 355-366.
<https://dl.acm.org/citation.cfm?id=948051>

Ng ICL (2014) *Creating New Markets in the Digital Economy: Value and Worth*. (Cambridge University Press, Cambridge).

Ng, ICL and Wakenshaw SYL (2017) Internet-of Things: Review and Research Directions. *International Journal of Research in Marketing*, 34(1): 3-21.

Ng ICL, Scharf K, Pogrebna G, Maull RS (2015) Contextual Variety, Internet-Of-Things and The Choice of Tailoring over Platform: Mass Customisation Strategy in Supply Chain Management. *International Journal of Production Economics*, 159:76-87.

- Nunan D, Di Domenico M (2013) Market Research and the Ethics of Big Data. *International Journal of Market Research*, 55: 505-520.
- Niederle M, Roth AE, Sonmez. T (2008) Matching and Market Design. Durlauf SN, Blume LE, eds. *The New Palgrave Dictionary of Economics*. 2nd Edition.
- Png I (1998) *Managerial Economics*. (Blackwell Publishers, Malden, MA).
- Poster M (2001) *What's the Matter With the Internet*. University of Minnesota Press, Minneapolis).
- Quah D (2003) Digital Goods and the New Economy. Chap. 13. Jones DC, ed. *New Economy Handbook* (Elsevier Academic Press, Amsterdam)
- Ramirez E, Brill J, Ohlhausen MK, Wright J, McSweeney T (2014) Data Brokers: A Call for Transparency and Accountability, Federal Trade Commission report
- Rose-Ackerman S (1985) Inalienability and the Theory of Property Rights. *Columbia Law Review*, 85(5): 931-969. doi:10.2307/1122458.
- Roth AE (1991) Game Theory as a Part of Empirical Economics. *The Economic Journal*, 101, No. 404 (Jan., 1991), pp. 107-114
- Salop SC (1979) Monopolistic competition with outside goods, *The Bell Journal of Economics*, 10(1): 141–156.
- Samuelson PA, Nordhaus WD (2004) *Economics*, 18th ed., McGraw Hill.
- Schwartz PM (2004) Property, Privacy, and Personal Data. *Harvard Law Review*, 117: 2056-2127.
- Shapiro C, Varian HR (1997) US government information policy. Unpublished manuscript, University of California, Berkeley.
<https://www.researchgate.net/publication/248244291>.
- Shapiro C, Varian HR (1998) *Information Rules: A Strategic Guide to the Network Economy*. (Harvard Business Press, Boston, MA)
- Spence M (1973) Job Market Signaling. *The Quarterly Journal of Economics*, 87(3): 35-74.
- Simon HA (2002) Near decomposability and the speed of evolution. *Industrial and Corporate Change*, 11(3): 587–599. <https://doi.org/10.1093/icc/11.3.587>
- Stiglitz JE (1975) The Theory of Screening, Education and the Distribution of Income. *American Economic Review*, 65(3): 283-300.

Swire PP, Litan RE (1998) *None Of Your Business: World Data Flows, Electronic Commerce, and the European Privacy Directive*. (Brookings Institution Press, Washington, D.C.)

Taylor CR (2004) Consumer Privacy and the Market for Customer Information. *RAND Journal of Economics*, 35 (4): 631-650.

TechAmerica Foundation's Federal Big Data Commission (2012) Demystifying Big Data: A Practical Guide to Transforming the Business of Government.
<http://www.techamerica.org/docs/fileManager.cfm?f=techamerica-bigdatareport-final.pdf>

Thisse J-F and Vives X (1988) On The Strategic Choice of Spatial Price Policy. *The American Economic Review*, 78(1): 122-137.

Tirole J (1988) *The Theory of Industrial Organization*. (MIT Press, Cambridge, MA)

Tuan J (2000) U.S. West, Inc. v. FCC, 15 Berkeley Tech. L.J. 353. Available at: <http://scholarship.law.berkeley.edu/btlj/vol15/iss1/18>

Transparency Market Research (2017) Data Broker Market, Global Industry Analysis, Size, Share, Growth, Trends and Forecast 2017 – 2026.
<https://www.transparencymarketresearch.com/data-brokers-market.html>

Van Zandt T (2004) Information Overload in a Network of Targeted Communication. *The RAND Journal of Economics*, 35(3): 542-560.

Vulkan N, Roth AE, Neeman Z (Eds) (2013) *The Handbook of Market Design*. (Oxford University Press, Oxford)

Wathieu L (2004) Consumer Habituation. *Management Science*, 50(5): 587-596.

Yoo Y, Henfridsson O and Lyytinen K (2010) The New Organizing Logic of Digital Innovation: An Agenda for Information Systems Research, *Information Systems Research*, 21(4), 724-735.